A UAV Based Automated Airborne Surveillance System

Michael P. Kontitsis, Nikos C. Tsourveloudis, and Kimon P. Valavanis

***Abstract* - This paper presents a system architecture for an integrated airborne surveillance system that allows Unmanned Aerial Vehicles (UAVs) to serve as autonomous sensing platforms. The proposed architecture consists of the communications, autonomous navigation and machine vision modules enabling the UAV to move into targeted locations, acquire and evaluate collected data while remaining in contact with the ground control station (GCS). The overall UAV system, although capable of operating fully autonomously, it also retains the option of semi-autonomous operation or even ground monitored teleoperation.**

***Index Terms*—Autonomous UAV, Machine vision, Surveillance.**

I. Introduction

THE role of airborne surveillance, whether autonomous, semi-autonomous or teleoperated, has been proved to be important and applicable to a wide range of applications such as search and rescue missions, border security, resource exploration, wildfire and oil spill detection. Until recently, common practice has been to use manned aircraft equipped with special sensors and to assign the actual recognition task (surveillance) to the crew or record image data and analyze them off-line on the ground [1]. In the exceptional cases where UAVs are used, they have been treated as sensor carrying platforms transmitting data to a GCS for analysis, since the involved UAVs lacked the ability of local on-board intelligence for data interpretation [2]. In addition, the absence of autonomous navigation couples tightly the UAV to the GCS.

The architecture presented in this paper addresses the UAV's dependency on the GCS for navigation and data interpretation, however it makes no assumption on how the UAV actually operates (autonomously or teleoperated). In essence, a navigation module and a machine vision system enable the UAV to fly over a specified (targeted) area while avoiding obstacles, acquire data and interpret them in real time, followed by decision making in terms of signaling an alarm or altering its course to get a better view of the targeted area. The whole process may be monitored from the GCS via the communication subsystem. The overall architecture of the automated surveillance system is shown in Fig.1, where the UAV may operate within the spectrum ranging from manually remote controlled system to a completely autonomous sensing agent.

As shown in Fig.1, there are two separate links to accommodate communication needs: the satellite link that provides required bandwidth for image transmission, and the RF link that is mainly used for flight parameter and control signal monitoring and adjustment. The two links are complementary to each other and either one may be used in case the other fails or it is unavailable.

The GCS receives data through the communication module that include the alert signals from the UAV. The flight is also monitored through the same module, and control is transferred completely to the GCS should it be necessary.

The UAV is guided to the specified observation area using autonomous navigation. The development of this subsystem is based on a neuro-fuzzy controller that constantly evaluates readings from the on board flight instruments and makes the appropriate adjustments to maneuver the UAV into desired position. The navigation system is also integrated with the machine vision system enabling the execution of predefined maneuvers to be triggered by the situation detected.

The integration of the navigation and communication module allows the ground operating crew to assert control of the flight at any time. The flight can be autonomous or semi-autonomous according to the requirements of each application or mission.

The rest of the paper is organized as follows. In the second section the Machine Vision Module is presented. Its description starts with an overview of the module, continues with the noise reduction, the feature extraction and classification subsystems and it concludes with some remarks on the computational complexity of the used algorithms. Also presented in section II, is an example of the module's function. Section III concludes the paper with some remarks on the issues that remain to be addressed in the future.

II. Machine Vision Module

A. *Overview*

The heart of the proposed automated surveillance system

This work was supported in part by the Greek Research and Technology Secretariat under Grant PAVET No. BE-00148.

M. P. Kontitsis is a Computer and Electronics Engineer and a M.Sc candidate at the Department of Production Engineering and Management, Technical University of Crete, 73100 Chania, GREECE (e-mail: konmix@dpem.tuc.gr).

N. C. Tsourveloudis (e-mail: nikost@dpem.tuc.gr) and K. P. Valavanis (e-mail: kimonv@dpem.tuc.gr) are with the Department of Production Engineering and Management, Technical University of Crete, 73100 Chania, GREECE.

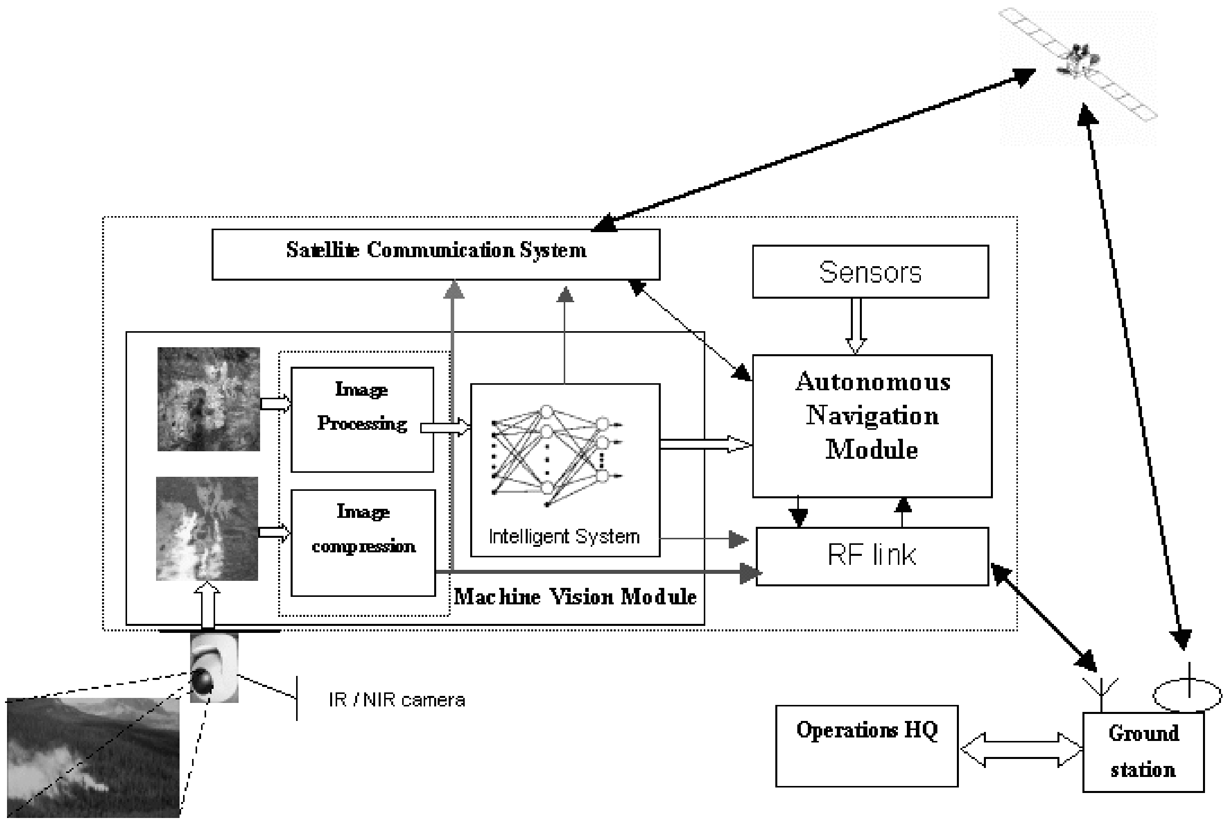

Fig.1: The architecture of the proposed integrated surveillance system

is the machine vision module. It consists of three components:

1. Noise reduction component
2. Feature extraction component
3. Feature vector classification component

The image is an 8-bit grayscale bitmap acquired by an Infrared (IR) or a Near Infrared NIR camera. It is subjected to preliminary image processing, namely gauss filtering, for noise reduction purposes. The *size* and the *mean intensity* of the various regions appearing on the image are then selected as image features as shown in Fig.2.

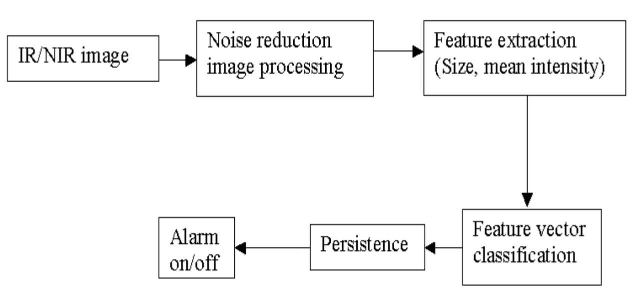

Fig.2: The block diagram of the machine vision module

The feature extraction from an image generates the feature vectors that are evaluated by the classifier. The classifier examines each feature vector and calculates the possibility of being the target defined by the application (e.g. a forest fire). If the assigned possibility of an area represented by a feature vector remains high for an adequate period of time, then the alarm is set on and the alarm signal is transmitted to the ground station.

Notice that due to the nature of the image sensors (IR/NIR camera) the data are a representation of the energy emitted by the various objects in the electromagnetic spectrum from 1μm to 14μm. These objects, especially those with a strong signature in the 3μm to 14μm band [3], are thermal sources that are likely to be the application objectives.

B. Noise Reduction

The image acquired by any camera is subject to noise. In order to avoid the undesirable side effects of noise such as misclassification, a spatial gaussian filter [4], [5] is used to suppress it. The use of this module is optional and depends heavily on the quality of the utilized camera.

C. Feature Extraction

A very important task in the whole surveillance process is to define the target that the system should look for and to select the features describing it in an almost unique manner. The region mean intensity as well as the region size are good candidates for features, because usually the regions of the image that correspond to the application's target have different intensity values than their surroundings. For example in an application regarding forest fire detection, targets that are trees on fire, are shown on the image as regions with higher mean intensity [3], which is defined as follows:

$$\text{Mean intensity} = \frac{\sum_{\forall \text{ pixel} \in \text{region}} grayvalue\_of\_pixel}{\text{number of pixels in the region}} \qquad (1)$$

Having chosen the vector features, all that remains is to extract them from the image. This is done using a region-growing algorithm [4] described below.

At first, points that will serve as seed points for the growing of the regions are selected. Such are all the pixels that are local maxima and exceed a certain high grayscale value (e.g. 190). This is based on the assumption that objects of interest appear brighter on the image. Having selected the seed points, the algorithm attempts to grow a region around each one of them by merging into the region any adjacent pixel whose intensity that does not deviate more than a given value from the intensity of the respective seed point. In other words the conditions for the annexation of a pixel p with intensity value g by a region S are:

$$|g - g(seed\ point)| < D \qquad (2)$$

p is 8-connected with a pixel belonging to the region S (3)

where $g(seed\ point)$ is the intensity value of the corresponding seed point and D is a constant.

This procedure is repeated until all pixels are assigned to some region. When this concludes, a series of feature vectors is created containing information for every region.

D. Feature Vector Classification

The development of the feature vector classifier is based on fuzzy logic. The choice of the fuzzy logic classifier is preferred for its simplicity and for the direct way in which it incorporates the experience into the structure of the classifier. It allows for easy changes on the system's behavior by fine-tuning the rule base or the membership functions that fuzzify the input variables, thus, making the system capable of adjusting to a variety of targets without much difficulty.

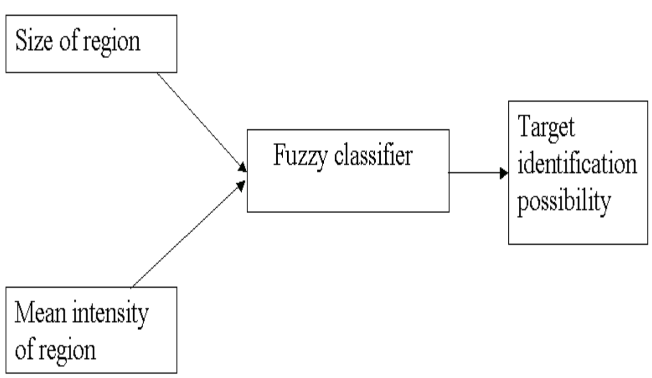

Fig.3: Block diagram of the feature vector classification subsystem.

The process of classifying the feature vectors is shown in Fig.3. At first the elements of each vector are assigned a membership function. This is done according to the nature of each feature and to specific target considerations. For example, as illustrated in Fig.4, the *mean intensity* feature is fuzzified to *Low*, *Mid* and *High* while referring only to the above 190 range of the grayscale values under the assumption that the target appears bright on the image. Recall that the image is an 8-bit grayscale so the intensity of the pixels takes 2^8 distinct values ranging from 0 (black) to 255 (white). Similarly, the size feature is divided to three linguistic variations *Small*, *Medium* and *Large*, respectively, as shown in Fig.5. Notice that the values on the x-axis depend heavily on the total number of pixels on the image and on the specific target. In this case the target is expected within the range of 20 to 200 pixels.

After completing the "fuzzification" process, the results are evaluated through a rule based inference engine. These rules are the ones that will determine the output of the system and therefore decide upon the presence or the absence of a target in the area. Rules are stated in the well known IF THEN form:

IF *Mean Intensity* is *High* AND *Size* is *Medium*
THEN *Target ID possibility* is *High.*

The output variable is the *target identification possibility* and its membership functions are depicted in Fig.6. After examining the rules and applying the aggregation method (max), the defuzzification process yields a number that classifies each region into one out of three different classes. If the output possibility is lower than 0.5 then the region, whose feature vector produced that result, has LOW possibility of being the designated target. A MEDIUM possibility classification occurs when the output is between 0.5 and 0.75, while HIGH possibility is assigned to those regions that produce an output higher than 0.75. The output of the system in respect to the input variables is shown in Fig. 7. A result of such a classification on a test image is presented in Fig. 8.

E. Persistence and Alarm Raising

Usually there are some random events, such as reflections or noise that can trigger an alarm by inducing into the image bright regions that do not actually correspond to a thermal source. Assuming that such events are small in duration, which is mostly the case, the frequency of such misclassifications may be reduced by introducing a sort of duration or time threshold. Under this condition, a region must produce a high possibility over several frames before it is considered as valid indication for the existence of the target.

When a region of the image is persistently classified as of a high possibility, then an alarm signal is transmitted both to the ground station through the communication module and directly to the autonomous navigation module. Also, images of the region are sent along with ordinary flight telemetry data.

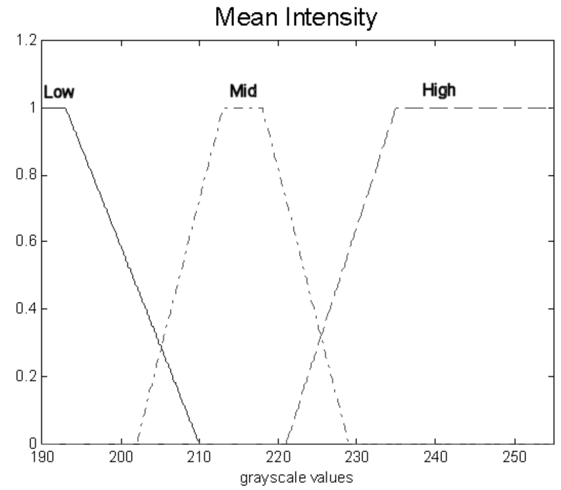

Fig.4: The Mean Intensity membership functions

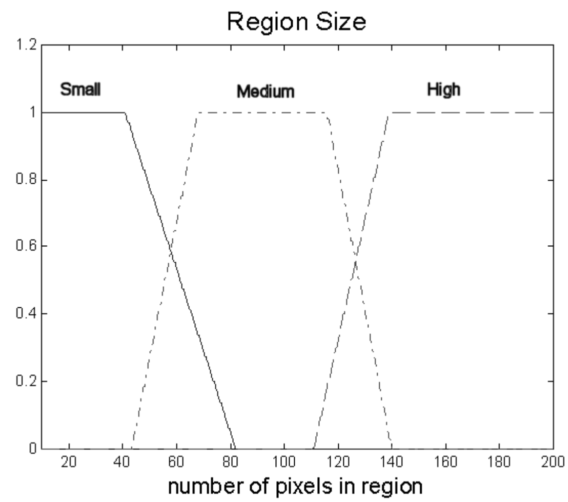

Fig.5: Membership functions of the Size of Region feature.

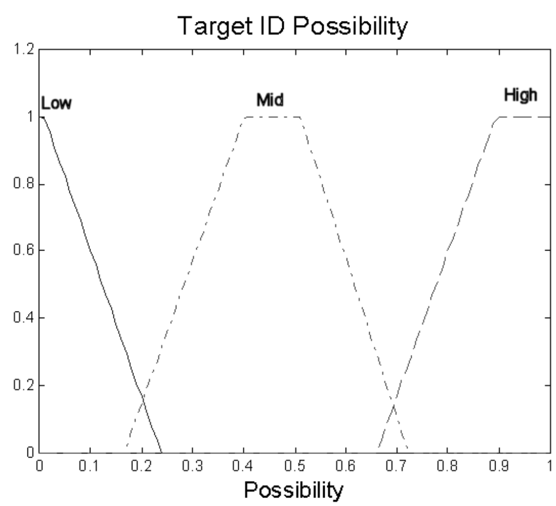

Fig.6: The membership functions for the output variable target identification possibility

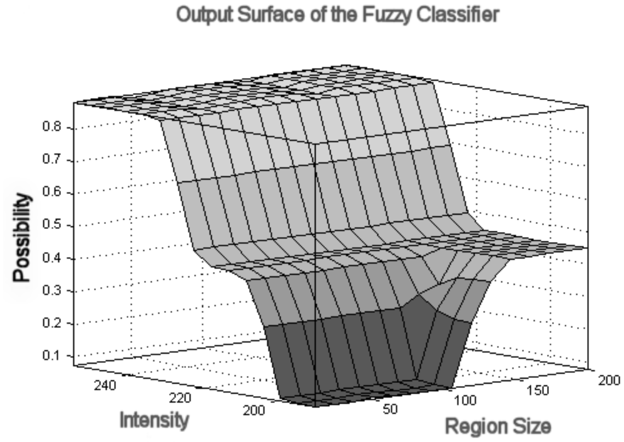

Fig.7: The output of the Fuzzy Classifier in respect to the input.

Along with the alarm signal, the coordinates of the thermal source that has been classified as the suspected target are computed. The coordinates of the thermal source are derived from the image using the well-known perspective projection model [4] [5] under which the coordinates of a point in a three dimensional space (*X, Y, Z)* is projected on the image plane at (*x, y*) as follows

$$x = f\frac{X}{Z} \quad , \tag{4}$$

$$y = f\frac{Y}{Z} \quad , \tag{5}$$

where *f* is the *focal length* of the camera's lens. Solving for *X* and *Y* yields respectively:

$$X = Z\frac{x}{f}, \tag{6}$$

$$Y = Z\frac{y}{f} \quad . \tag{7}$$

Given that *x* and *y* are the image's plane coordinates, the actual 3D coordinates of the suspected target could be calculated, by using equations (6) and (7), assuming that there is some information about *Z*. This information is either provided by a laser scanner mounted on the UAV or may be extracted from a digital elevation map of the region that the UAV is flying above. In the absence or failure of the above, a rough estimation is possible assuming that:

1. Every point visible to the camera is about as far from the camera as any other
2. The distance from the ground can be approximated by the UAV's altimeter.

These assumptions are valid for a narrow Field of View lens and an UAV flight close to the sea level. When they hold, the *Z* coordinate of every point projected on the image plane is approximated by the value of the altitude.

The preceding analysis is further based on the assumption that the world coordinates system coincides with the camera coordinates system. As a result of that, the position of the region is estimated relatively to the UAV. Further localization on a map is possible using the information about the UAV's position provided by the altimeter and the on board GPS.

F. Image Compression

The presentation of this subsystem has been intentionally left last because it does not participate in the process of data interpretation. Its presence is solely dictated by reasons regarding the efficient use of the communication channel. So, the images that the ground station receives are compressed versions of the ones processed onboard the UAV.

G. Complexity

Dealing with two-dimensional images, may be computationally intensive. However, in this case, the algorithms used by the machine vision module, namely the noise reduction and region growing, did not exceed the order of $O(n^2)$ in complexity, with n being the dimension of a square image (nxn). Also, under the current implementation, the fuzzy classifier is found to have its own complexity depending mainly on the product of the number of features (n) and the number of rules in its inference engine (m). So, it can be said that it is of order $O(n\text{x}m)$.

Although one can argue that $O(n^2)$ algorithms do not scale nicely, the various tests, that the system has been subjected to, showed that, for images of resolution up to 320x200 pixels, real time or near real time performance can be achieved on average commodity hardware. This resolution is found on the most mid valued IR cameras available today and it is sufficient for most applications.

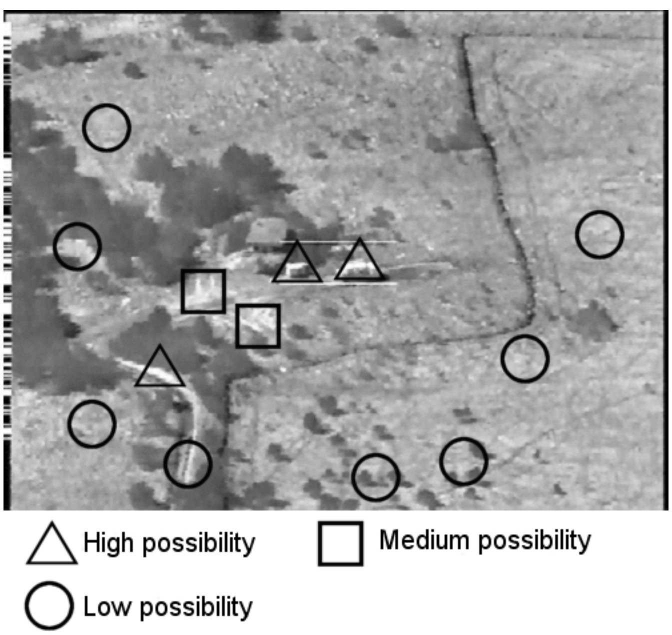

Fig.8: A sample image showing the classification of various regions.

III. Concluding Remarks

The development of a platform under the architecture presented here is at experimental stage. Several issues remain to be addressed in practice such as low level interfacing between the modules. Also the intelligent system that has the role of identifying targets must be enhanced to include a learning mechanism, which will allow it to gain experience and perform more accurate recognition over time.

IV. References

[1] G.G. Wright, K.B. Matthews, W.M. Cadell and R. Milne, "Reducing the cost of multi-spectral remote sensing: combining near-infrared video imagery with colour aerial photography", *Computers and Electronics in Agriculture*, Vol 38, Issue 3, March 2003, pp. 175-198.

[2] Akira Sato, "Research, development and civil application of an autonomous, unmanned helicopter", Aeronautic Operations, YAMAHA MOTOR CO., LTD., Shizuoka, Japan.

[3] Floyd F. Sabins, Jr, *Remote Sensing Principles and Interpretation,* Freeman, New York,. 1990.

[4] R. Gonzalez and R. Woods, *Digital Image Processing*, Addison Wesley, 1992.

[5] Ramesh Jain, Rangachar Kasturi and Brian G. Schunck, *Machine Vision,* Mc Graw Hill, 1995.

[6] S. Pal, S. Mitra, *Neuro-fuzzy pattern recognition: methods in soft computing*, Wiley, New York, 1999.

[7] Jun Xu, Jiang-qi Zhang and Chang-hong Liang, "Prediction of the performance of an algorithm for the detection of small targets in infrared images", *Infrared Physics & Technology*, Vol. 42, 2001, pp. 17-22.

[8] James C. Tilton, "Image Segmentation by Region Growing and Spectral Clustering with a Natural Convergence Criterion", NASA's Goddard Space Flight Center, *Proceedings of the 1998 International Geoscience and Remote Sensing Symposium (IGARSS '98)*, Seattle, WA, July 6-10, 1998.